Application Serial No. 09/844,120

REMARKS

The present application was filed on April 27, 2001 with claims 1-24. In the outstanding Office Action dated February 16, 2005, the Examiner has: (i) rejected claims 1-3, 6, 10-12, 18, 20, 21 and 23 under 35 U.S.C. §102(b) as being anticipated by U.S. Patent No. 5,880,788 to Bregler (hereinafter "Bregler"); (ii) rejected claims 4 and 5 under 35 U.S.C. §103(a) as being unpatentable over Bregler, in view of U.S. Patent No. 6,256,046 to Waters et al. (hereinafter "Waters"); (iii) rejected claims 7 and 8 under §103(a) as being unpatentable over Bregler, in view of U.S. Patent No. 6,580,437 to Liou et al. (hereinafter "Liou"); (iv) rejected claims 9 and 22 under §103(a) as being unpatentable over Bregler in view of U.S. Patent No. 6,250,928 to Poggio et al. (hereinafter "Poggio"), and further in view of Liou; (v) rejected claims 13-15, 17, 19 and 24 under §103(a) as being unpatentable over Bregler, in view of U.S. Patent No. 5,884,267 to Goldenthal et al. (hereinafter "Goldenthal"); and (vi) rejected claim 16 under §103(a) as being unpatentable over Bregler in view of Goldenthal, and further in view of Waters.

In this response, claims 1, 2, 10, 11, 13, 18, 21 and 24 have been amended. Applicants traverse the §102(b) and §103(a) rejections for at least the reasons set forth below. Applicants respectfully request reconsideration of the present application in view of the above amendments and the following remarks.

Claims 1-3, 6, 10-12, 18, 20, 21 and 23 stand rejected under §102(b) as being anticipated by Bregler. The Examiner merely maintains the rejections set forth in his prior Office Action dated June 25, 2004, contending that Bregler discloses all of the elements set forth in the subject claims. Applicants respectfully disagree with this contention. Independent claims 1, 10 and 18, which are of similar scope, require capturing images of body movements corresponding to one or more words in an utterance and presenting each image segment with a corresponding decoded text word in that utterance. It is to be emphasized that, unlike Bregler, the claimed invention does not utilize images from a previously recorded video sequence and attempt to match arbitrary utterances with the prerecorded images. Rather, the captured image segments used by the claimed invention are generated directly from the one or more words in the utterance and, as such, may be considered analogous to the spectral feature vector set generated by the ASR engine, which are then sent to an image player for presenting each image segment with the corresponding decoded word. While

8

Application Serial No. 09/844,120

Bregler may be capable of synchronizing an <u>arbitrary utterance</u> with a prerecorded image sequence of lip movements stored in a database (Bregler; column 2, lines 35-36), since Bregler uses prerecorded images that are <u>not generated from the actual utterance being presented</u> with the images, Bregler fails to provide multiple sources of information for comprehending and/or verifying the accuracy of the decoded speech, and is thus directed to an entirely different problem than that solved by the present invention.

In response to Applicants' arguments distinguishing the claimed invention from the cited prior art, the Examiner states that "the image-capturing step in the base claims 1, 10, and 18 <u>does not</u> mention anything about capturing <u>live</u> images corresponding to one or more words in the utterance. The process of capturing images in the base claims 1, 10, and 18, can capture either prerecorded or live images as long as there are images available at the input of the claimed visual detector" (final Office Action; page 2, paragraph 1; emphasis in original). In this regard, the Examiner's statement does not address the primary argument that Bregler simply fails to disclose capturing images of body movements corresponding to one or more words in the utterance, as explicitly required by the claimed invention. Furthermore, the Examiner contends that Bregler "discloses a process of analyzing images to identify visual features such as speaker's lip position (col. 5, lines 49-60)" (final Office Action; page 2, paragraph 1). However, Applicants assert that any images which Bregler analyzes to identify such visual features are not images generated from the actual words in the utterance to be presented, and thus Bregler is clearly distinguishable from the claimed invention.

As stated in Applicants' prior response dated September 27, 2004, while Bregler may disclose, with reference to FIG. 1, analyzing a stored <u>video recording</u> of a person speaking at step S1 "to associate characteristic sounds in the utterance with specific video image sequences," and locating, at step S2, "salient features" in the image sequence (Bregler; column 4, lines 1-6), Bregler fails to disclose capturing images relating to the actual utterance to be presented, as required by claims 1, 10 and 18. In fact, Bregler specifically states (*see. e.g.*, Bregler; column 4, lines 24-25) that a video stream is produced of a person speaking a <u>new</u> utterance (i.e., <u>other than the original utterance</u> from which the prerecorded video was extracted) by synchronizing a <u>stock video recording</u> to a new soundtrack (Bregler; column 3, lines 66-67). Thus, Bregler merely discloses matching an

9

Application Serial No. 09/844,120

arbitrary utterance to be presented with prerecorded images, "without requiring a video recording of the new sounds being uttered" (Bregler; column 2, lines 19-22). Bregler fails to disclose capturing images corresponding to one or more words in the utterance being presented, as required by the subject claims.

Notwithstanding the above traversal, however, independent claims 1, 10 and 18 have amended in order to provide further clarity. These amendments are not believed to require further consideration and/or search, and therefore entry of the amendments made herein is respectfully requested. Specifically, claims 1, 10 and 18, as amended, recite that the visual detector captures "images of body movements substantially concurrently from the one or more words in the utterance." Support for this amendment may be found in the present specification, for example, beginning on page 5, line 25, where it states:

> the visual feature extractor 102 preferably includes an image detector 110, such as, for example, a digital or video camera, charge-coupled device (CCD), or other suitable alternative thereof, for capturing images or clips (i.e., a series of successive images in time) of lip movements, sampled at one or more predetermined time intervals, generated by a given speech utterance. (emphasis added)

The above-noted amendment is intended to address the Examiner's contention that capturing live images corresponding to one or more words in the utterance is not present in the subject claims. The prior art of record fails to teach or suggest at least this feature of the claimed invention.

Likewise, Bregler fails to disclose an image player configured for "receiving and presenting the decoded word with each image segment generated therefrom," as recited in amended claim 1, wherein the decoded word refers to decoded speech text from the ASR system. While Bregler may disclose image analysis in the general sense at step S2 based on prerecorded video image sequences, Bregler clearly fails to disclose extracting features from the captured image sequence based on the actual utterance to be presented, as required by the claimed invention. Because Bregler does not utilize captured images from the actual utterance to be presented, Bregler is simply not capable of achieving an important objective of the claimed invention, namely, comparing and verifying the decoded speech text with the images obtained therefrom in order to determine the accuracy of the recognized text (see, e.g., Specification; page 4, lines 11-17).

10

Application Serial No. 09/844,120

With regard to the image player recited in claim 1, the Examiner contends that Bregler discloses "an image player operatively coupled to the visual feature extractor, the image player receiving and presenting each image segment with the corresponding decoded word (figures 6-7, the decoded word can be presented audibly)" (final Office Action; page 5, paragraph 10). Applicants respectfully disagree with this contention and submit that Bregler fails to specifically disclose that the image segments are displayed with corresponding decoded speech text, as required by the subject claims. Moreover, even if Bregler teaches, with reference to figures 6-7, presenting the decoded word audibly, as the Examiner contends, Applicants assert that this is entirely distinguishable from amended claim 1, which explicitly requires that the image player present "the decoded word with each image segment generated therefrom." The prior art of record fails to disclose at least this feature of the claimed invention.

For at least the above reasons, Applicants assert that claims 1, 10 and 18 are patentable over the prior art. Accordingly, favorable reconsideration and allowance of these claims are respectfully solicited.

With regard to claims 2, 3, and 6, which depend from claim 1, claims 11 and 12, which depend from claim 10, and claims 20, 21 and 23, which depend from claim 18, Applicants submit that these claims are also patentable over the prior art of record by virtue of their dependency from their respective base claims, which are believed to be patentable for at least the reasons set forth above. Moreover, one or more of these claims define additional patentable subject matter in their own right. For example, claims 2 and 11, as amended, further define the image player as being configurable for repeatedly presenting one or more image segments with the corresponding decoded word "by looping on a time sequence of successive images corresponding to the decoded word" (emphasis added). In response to Applicants' arguments, the Examiner contends that "the features upon which applicant relies (i.e., "repeatedly presenting" refers to "looping on a time sequence of successive images associated with a particular word(s) in the utterance" ...) are not recited in the rejected claim(s) (final Office Action; page 2, paragraph 2; emphasis in original). The amendments to claims 2 and 11 presented herein are therefore intended to clarify the definition of the term "repeatedly presenting" set forth in claims 2 and 11, as suggested by the Examiner.

11

Application Serial No. 09/844,120

Likewise, claims 3 and 12 further define the apparatus as including a delay controller for "selectively controlling a delay between an image segment and a corresponding decoded word" in the utterance. Bregler fails to disclose at least this additional feature of the claimed invention. In response to Applicants prior arguments relating to claims 3 and 12, the Examiner contends that aligning an image segment with a corresponding decoded word must involve time controlling, and that time controlling involves adjusting a time delay in order to realize synchronization (time warping) (final Office Action; page 3, paragraph 3). Applicants respectfully disagree with this contention. First, "time warping" as taught by Bregler merely relates to aligning a new soundtract to a prerecorded image sequence. The prerecorded image sequence is not generated from the new soundtrack, as is required in claims 3 and 12. Furthermore, as previously stated, although Bregler may disclose the use of "time warping" to align a new soundtrack to a prerecorded image sequence, Bregler defines time warping as dropping one or more frames from the original recording, "so that the remaining frames in a new video sequence 27 correspond to the timing of the new speech soundtrack 20" (Bregler; column 10, lines 25-30). The concept of time warping taught by Bregler does not involve controlling a delay at all, and is thus distinguishable from the delay controller recited in claims 3 and 12.

For at least the reasons set forth above, claims 2, 3, 6, 11, 12, 20, 21 and 23 are believed to be patentable over the prior art of record, not merely by virtue of their dependency from their respective base claims, but also in their own right. Accordingly, favorable reconsideration and allowance of claims 2, 3, 6, 11, 12, 20, 21 and 23 are respectfully requested.

Claims 4 and 5 stand rejected under §103(a) as being unpatentable over Bregler, in view of Waters. The examiner acknowledges that "Bregler does not disclose a position detector coupled to the visual detector, the position detector comparing the position of the user with a reference position and generating a control signal . . . and a label generator coupled to the position detector" (final Office Action; page 9, paragraph 21). However, the Examiner contends that such features are disclosed in Waters, particularly at column 4, lines 20-41 and column 5, lines 28-59 (final Office Action; page 9, paragraph 21). While Applicants respectfully disagree with this contention, Applicants assert that claims 4 and 5, which depend from claim 1, are patentable over the prior art of record by virtue of their dependency from claim 1, which is believed to be patentable for at least

12

Application Serial No. 09/844,120

the reasons set forth above. Accordingly, favorable reconsideration and allowance of claims 4 and 5 are respectfully solicited.

Claims 7 and 8 stand rejected under §103(a) as being unpatentable over Bregler, in view of Liou. The Examiner acknowledges that "Bregler fails to specifically disclose that the image segments are displayed with corresponding decoded speech text"(final Office Action; page 10, paragraph 25). However, the Examiner contends that such a feature is disclosed in Liou (final Office Action; page 11, paragraph 25). In response to Applicants arguments relating to claims 7 and 8, the Examiner contends that Bregler and Liou are combinable, with Liou being relied upon for teaching the displaying of decoded speech text on the display (final Office Action; page 3, paragraph 5). While Applicants maintain their assertion that Bregler and Liou are not analogous art, and are therefore not believed to be combinable (there is various case law regarding the strict requirement for a motivation to combine references), Applicants submit that Bregler and Liou, when considered in combination, fail to disclose all of the elements set forth in claims 7 and 8.

Applicants submit that claims 7 and 8, which depend from claim 1, are patentable over the prior art of record by virtue of their dependency from claim 1, which is believed to be patentable for at least the reasons set forth above. Moreover, these claims define additional patentable subject matter in their own right. For example, claim 7 further defines the apparatus as including a display controller configured to selectively control "one or more characteristics of <u>a manner in which the image segments are displayed</u> with corresponding decoded speech text" (emphasis added). The prior art of record fails to teach or suggest at least this additional feature of the claimed invention. In this regard, the Examiner contends that Bregler teaches controlling a manner in which image segments are displayed with the corresponding audio (at col. 10, lines 13-30) (final Office Action; page 10, paragraph 25). Applicants respectfully disagree with this contention.

Specifically, Bregler fails to disclose selectively controlling a manner in which <u>decoded words</u> from an ASR are displayed with captured image segments corresponding thereto, as required by the subject claims. Liou fails to disclose selectively controlling <u>a manner</u> in which closed-captioned text is displayed with video programming, and thus fails to supplement the deficiencies of Bregler. Claims 7 and 8 are therefore believed to be patentable over the combination of Bregler

13

Application Serial No. 09/844,120

and Liou, not merely by virtue of their dependency from claim 1, but also in their own right. Accordingly, favorable reconsideration and allowance of claims 7 and 8 are respectfully requested.

Claims 9 and 22, which depend from claims 1 and 18, respectively, stand rejected under §103(a) as being unpatentable over Bregler in view of Poggio, and further in view of Liou. While Applicants maintain the assertion that Bregler, Poggio and Liou are not analogous art, and are therefore not believed to be combinable, Applicants submit that Bregler, Poggio and Liou, when considered in combination, fail to disclose all of the elements set forth in claims 9 and 22. Specifically, contrary to the Examiner's contention in this regard, Poggio fails to disclose an image player displaying "each image segment in a separate window on a display in close proximity to the decoded speech text corresponding to the image segment," as required by claims 9 and 22. The Examiner relies on the disclosure in FIG. 8 and at column 10, lines 52-67 of Poggio as support for such teaching. However, no such teaching exists. Rather, Poggio, in FIG. 8, merely illustrates how a new visual utterance is synthesized from respective visemes. The individual visemes are not displayed to a viewer in separate windows, but "are concatenated, or put together and played seamlessly one right after the other" as part of a video sequence in a single display window (Poggio; column 11, lines 1-4; emphasis added). Therefore, claims 9 and 22 are believed to be patentable over the prior art of record, not merely by virtue of their dependency from their respective base claims, but also in their own right. Accordingly, favorable reconsideration and allowance of claims 9 and 22 are respectfully solicited.

Claims 13-15, 17, 19 and 24 stand rejected under §103(a) as being unpatentable over Bregler, in view of Goldenthal. With regard to independent claims 13 and 24, which are of similar scope, the Examiner contends that Bregler and Goldenthal, when considered in combination, disclose all of the elements set forth in these claims. Applicants respectfully disagree with this contention. As previously explained, Bregler merely attempts to match prerecorded video images with words in an arbitrary utterance. To do this, Bregler employs a database of stored image clips and corresponding words. The stored video images are not captured from the utterance being presented. Moreover, Goldenthal fails to supplement the deficiencies of Bregler in this regard. Furthermore, while Goldenthal may disclose that acoustic-phonetic units are formatted as data records including a "starting time 231, an ending time 232, and an identification 233 of the corresponding acoustic-

14

Application Serial No. 09/844,120

phonetic unit (Goldenthal; column 4, lines 14-18), neither Bregler nor Goldenthal teaches or suggests "aligning the plurality of images into one or more image segments according to the start and stop times received from the ASR system, wherein each image segment corresponds to a decoded word in the utterance," as required by claims 13 and 24.

Notwithstanding the above traversal, however, claims 13 and 24 have been amended to provide further clarification. Specifically, claim 13, as well as claim 24 which is of similar scope, as amended, recite "capturing a plurality of images representing body movements substantially concurrently from the one or more words in the utterance; associating each of the captured images generated from the one or more words in the utterance with time information relating to an occurrence of the image; . . . and presenting the decoded word with the corresponding image segment generated therefrom." Applicants submit that the prior art of record fails to teach or suggest such features of the subject claims.

As previously stated, Bregler may disclose image analysis in the general sense at step S2 based on prerecorded video image sequences, but Bregler fails to disclose extracting features from the captured image sequence based on the actual utterance to be presented. In Goldenthal, starting times and ending times associated with each data record are used to translate the acoustic-phonetic units into visemes by a rendering subsystem (Goldenthal; column 4, lines 20-22). However, the rendering subsystem, like Bregler, does not utilize captured images corresponding to the utterance to be presented. Goldenthal fails to supplement the deficiencies of Bregler, and therefore the combination of Bregler and Goldenthal fails to teach or suggest all of the limitation set forth in amended claims 13 and 14.

For at least the reasons given above, Applicants assert that claims 13 and 24 are patentable over the prior art. Accordingly, favorable reconsideration and allowance of these claims are respectfully requested.

With regard to claims 14, 15 and 17, which depend from claim 13, and claim 19, which depends from claim 18, Applicants submit that these claims are also patentable over the prior art of record by virtue of their dependency from their respective base claims, which are believed to be patentable for at least the reasons set forth above. Moreover, one or more of these claims define additional patentable subject matter in their own right. For example, claim 14 further defines the

15

Application Serial No. 09/844,120

method as including the step of "selectively controlling a delay between when an image segment is presented and when a decoded word corresponding to the image segment is presented." Likewise, claim 15 further defines the method as including the step of "selectively controlling a manner in which an image segment is presented with a corresponding decoded word." The prior art of record fails to teach or suggest at least these additional features of the claimed invention.

For at least the reasons set forth above, claims 14, 15, 17 and 19 are believed to be patentable over the prior art of record, not merely by virtue of their dependency from their respective base claims, but also in their own right. Accordingly, favorable reconsideration and allowance of claims 14, 15, 17 and 19 are respectfully solicited.

Lastly, claim 16 stands rejected under §103(a) as being unpatentable over Bregler in view of Goldenthal, and further in view of Waters. The Examiner contends that the combination of Bregler, Goldenthal and Waters discloses all of the elements set forth in claim 16. While Applicants respectfully disagree with the Examiner's contention in this regard, Applicants submit that claim 16, which depends from claim 13, is also patentable over the cited prior art by virtue of its dependency from claim 13. Accordingly, favorable reconsideration and allowance of claim16 is respectfully requested.

In view of the foregoing, Applicants believe that pending claims 1-24 are in condition for allowance, and respectfully request withdrawal of the §102 and §103 rejections.

Respectfully submitted,

Date: May 13, 2005

Wayne L. Ellenbogen
Attorney for Applicant(s)
Reg. No. 43,602
Ryan, Mason & Lewis, LLP
90 Forest Avenue
Locust Valley, NY 11560
(516) 759-7662

16